Keep it simple: estimation strategies for ordered response models with fixed effects

Maximilian Riedl*and Ingo Geishecker

Georg-August-Universität Göttingen, Faculty of Economic Sciences

Version: September 2012

Abstract

By running Monte Carlo simulations, we compare different estimation strategies of ordered response models in the presence of non-random unobserved heterogeneity. We find that very simple binary recoding schemes deliver unbiased and efficient parameter estimates. Furthermore, if the researcher is interested in the relative size of parameters the simple linear fixed effects model is the method of choice.

Keywords: fixed effects ordered logit, ordered responses, happiness

JEL classifications: C230, C250, I310

*Corresponding Author: Maximilian Riedl, Georg-August-Universität Göttingen, Platz der Göttinger Sieben 3, 37073 Göttingen, Germany, e-mail: maximilian.riedl@wiwi.uni-goettingen.de, Tel: +49 551 394878, Fax: +49 551 397880. The authors are grateful to Stefan Sperlich, Paul Frijters and Gregori Baetschmann for valuable comments and Martin Breßlein for programming support.

1 Introduction

When estimating models for longitudinal ordinal response data, researchers typically face the problem of accounting for unobserved personality traits that may be correlated with explanatory variables, while at the same time accommodating the ordinal nature of the dependent variable. Since there is no consistent estimator for an ordered logit or probit model that can explicitly incorporate individual fixed effects, different estimation strategies have been pursued in the literature. Yet, the literature provides no guideline for when to use which estimator.

Authors such as Winkelmann and Winkelmann (1998), Senik (2004), Clark (2003) and Kassenböhmer and Haisken-DeNew (2009) recode the ordinal dependent variable into a binary variable and subsequently apply the conditional logit estimator of Chamberlain (1980). This approach has the advantage that it maintains the nonlinear character of the dependent variable. However, recoding ordinal responses into binary responses requires the researcher to more or less arbitrarily define a threshold above which the dependent binary variable takes the value one. As a consequence, potentially important variation in the original ordinal response variable is disregarded.

Extending this approach, Ferrer-i-Carbonell and Frijters (2004) propose an estimation strategy that uses much more of the variation in the ordinal response variable for binary recoding. However, since this procedure requires calculation of the individual Hessian for each binary recoding option, it is computationally very expensive. Nevertheless, the estimator has gained some popularity and has been employed in a number of recent empirical studies, such as Frijters et al. (2006), Frijters et al. (2004), Knabe and Rätzel (2009), Clark et al. (2010) and Geishecker et al. (2012).

Another binary recoding strategy is developed in Baetschmann et al. (2011). Their so called "Blow-Up and Cluster" (BUC) estimator aims at using all variation of the ordinal response variable by expanding the data set to accommodate all possible binary recoding options of the ordered dependent variable. The approach has been used in, e.g., Geishecker et al. (2012).

A fourth and very common approach taken, for example, by Di Tella et al. (2001), Scheve and Slaughter (2004), and Senik (2004), assumes cardinality of the ordered response variable and estimates a simple first difference or within-transformed linear model. Although certain applications, such as studies of subjective well-being, have shown that the cardinality assumption does not severely bias estimates (see Ferrer-i-Carbonell and Frijters, 2004), it is difficult to generalize this finding to other applications. To circumvent violations of the cardinality assumption van Praag and Ferrer-i-Carbonell (2008) propose to rescale the ordered dependent variable to a normal distributed variable centered around zero. The so called "probit-adapted OLS" technique has been used by, e.g., Stevenson and Wolfers (2008), Luechinger (2009), Clark et al. (2010), Luechinger et al. (2010), and Geishecker (2012).

Choosing from this arsenal of estimation strategies is not an easy task, since apart from rough comparisons of the alternatives discussed in the context of concrete applications (e.g. Ferrer-i-Carbonell and Frijters, 2004), there is little comparative evidence on their finite sample properties and performance that can be generalized. In the present paper, we aim to fill this gap by performing Monte Carlo simulations that yield statistical measures for consistency and efficiency for the previously mentioned alternative estimation strategies.

The contribution of the paper is twofold. First, the paper presents a systematic evaluation of the recently developed conditional binary estimators for ordered response models in finite samples, which are unknown so far. Second, the paper functions as a guide for applied researchers who typically face data for which asymptotic theory is not applicable and who need to choose between the different proposed estimation strategies.

The remainder of the paper is structured as follows: Section 2 revisits the proposed estimation strategies more formally. Section 3 describes the Monte Carlo experiment, including the data generating process, and presents the results of our simulations for different sample sizes, ordinal scales, number and distribution of covariates. Section 4 concludes.

2 Estimation Strategies in Detail

We want to estimate a latent variable model with ordered response data. The model is given by:

$$y_{it}^* = \beta' x_{it} + \alpha_i + \epsilon_{it} \tag{1}$$

where y_{it}^*, for example, represents general well-being of individual $i = 1, \ldots, I$ at time $t = 1, \ldots, T$ and is a continuous variable that cannot be observed. x_{it} is a vector of independent explanatory variables, α_i is the individual personality trait assumed to be correlated with the vector of explanatory variables x_{it}. Finally ϵ_{it} is the logistically distributed error term. Since the continuous latent variable y_{it}^* cannot be observed, an ordered categorical response variable y_{it} is measured with $k = 1, \ldots, K$ categories and individual-specific thresholds λ_k^i, where $\lambda_k^i < \lambda_{k+1}^i$:

$$y_{it} = k \Leftrightarrow \lambda_k^i \leq y_{it}^* < \lambda_{k+1}^i. \tag{2}$$

In what follows we discuss and compare six possible estimation strategies for this ordered response problem. One simple estimation strategy for ordered response data with unobserved personality traits is to transform the ordered response variable so that it can be estimated with a conditional logit estimator (see Chamberlain, 1980). To generate the required binary response variable from ordered responses one common approach is to apply what is considered a meaningful threshold (Y) to the whole data set (e.g., Winkelmann and Winkelmann, 1998; Clark, 2003) such that:

$$B_{it} = \begin{cases} 0 & \text{if } y_{it} \leq Y \\ 1 & \text{if } y_{it} > Y. \end{cases} \tag{3}$$

The conditional logit statistic corresponding to this simple coding scheme then is:

$$P\left[B_{it} | \sum_t B_{it} = c_i\right] = \frac{e^{\sum_{t=1}^T B_{it} x_{it} \beta}}{\sum_{y \in S(k_i, c_i)} e^{\sum_{t=1}^T B_{it} x_{it} \beta}} \quad . \tag{4}$$

This represents the probability that the dependent variable is above Y, conditional on the sum c_i. More precisely, c_i denotes the number of times the dependent variable per group exceeds the threshold Y, $0 < c < T$. S describes the set of all possible combinations of $y_{i1}, \ldots, y_{iT}$ that sum up to $\sum_t B_{it} = c_i$. In the following, we refer to this estimation strategy as simple conditional logit (SCLOG).

Clearly the SCLOG ignores all variation in y_{it} that takes place below or above Y. Furthermore and most importantly, the applied simple coding scheme also abstracts from the possibility that the thresholds λ_k^i in equation 2 vary in i. For example, consider ordered responses on life satisfaction. Our sample may include a happy life long enthusiast and an equally happy life-long sceptic. While the enthusiast's self reported life satisfaction scores may tend to be on the high side, responses of the equally happy sceptic may tend to be on the low side. Accordingly, in this example, a common threshold crossing cannot capture changes in the self-reported life satisfaction of the sceptic and the enthusiast equally well. Thus, this strategy does not address personality traits in any satisfactory way.

A somewhat more sophisticated coding scheme takes account of such personality traits by constructing a binary response variable (E) that takes the value one if the score of the ordered categorical response variable is above the individual-specific mean of all ordered categorical responses:

$$E_{it} = \begin{cases} 0 & \text{if } y_{it} \leq E(y_{it}) \\ 1 & \text{if } y_{it} > E(y_{it}) \quad . \end{cases} \tag{5}$$

To stay with the example, our enthusiast and sceptic now have different thresholds that reflect that the responses of the former tend to be on the high side of the ordered scale while the responses of the latter tend to be on the low side. Recent applications of this approach include Kassenböhmer and Haisken-DeNew (2009). In the following, we refer to this approach as individual mean conditional logit (IMCLOG).

An extension to the IMCLOG method is proposed in Ferrer-i-Carbonell and Frijters (2004) taking into account more variation in individuals' ordered responses. Their method uses the conditional logit approach combined with a fairly complex individual-specific coding of the dependent variable. They use the information from the second derivative of the log likelihood function, the Hessian matrix, per individual to choose which coding is appropriate for the final conditional logit estimation. This procedure consists of three steps, which deserve some detailed explanation as the exposition in the original article of Ferrer-i-Carbonell and Frijters (2004) is incomplete.

In the first step the ordered dependent variable y_{it} with K categories is split into $K-1$ new binary coded variables D_{ik} capturing all possible threshold crossings.

The first newly generated variable D_{i1} equals one if the original dependent variable y_{it} is at least one category greater than the minimum of y_{it} for each i:

$$D_{itk} = \begin{cases} 0 & \text{if } y_{it} \leq min_i\{y_{it}\} \\ 1 & \text{if } y_{it} > min_i\{y_{it}\} \end{cases} \tag{6}$$

The next newly generated variable D_{i2} equals one if the original dependent variable is at least two categories greater than the minimum of y_{it} for each i and so forth. A more detailed example can be found in the appendix of Ferrer-i-Carbonell and Frijters (2004).

In a second step, a conditional logit model (Chamberlain, 1980) is estimated for the first threshold crossing to derive the coefficients (β) that are used to calculate the Hessian matrix for each individual for each D_{ik}.

The first and second derivatives of the log likelihood function used for these calculations can be found in the appendix to this paper. On this basis, the sum of the diagonal elements, the so called "trace," for each individual Hessian is calculated for each D_{ik}. The final binary dependent variable is then generated by choosing the specific D_{ik} that corresponds to the minimum trace per individual i. Since the variance of the estimated conditional logit coefficient is the negative of the inverse of the sum of the Hessian H_i over all i, this yields the maximum likelihood estimator with minimal variance.

In a third step, the newly generated binary variable, which reflects the optimal choice of D_{ik} for all i, is fed into a conditional logit estimation to obtain the final coefficients. In the following, we refer to this estimation strategy as the Ferrer-i-

Carbonell Frijters estimator (FCF). Since the FCF estimator requires calculation of individual-specific Hessian matrices for each possible threshold D_{ik}, it is computationally expensive, particularly if T is large.[1]

Note that the individual-specific coding procedure based on minimum-trace individual Hessian matrices is initially based on the assumption of knowing the true parameter estimates of the latent variable model. It is debatable how these initial parameters should be obtained. We test whether the FCF estimation results differ when using the individual mean coding procedure (IMCLOG), i.e., whether the FCF estimates are sensitive to replacing D_{it1} with E_{it} from Equation 5. Furthermore, we also estimate an iterated version of the FCF, continuously updating the initial parameters. However, there are only subtle differences between the corresponding final FCF parameters. Thus, the FCF method is robust with respect to the choice of the first-step estimation routine.

Yet, an alternative recoding scheme is introduced in Baetschmann et al. (2011). Their so called "Blow-Up and Cluster" (BUC) estimator recodes the original dependent variable with k categories into $k-1$ different dichotomizations using $k-1$ different thresholds. Each observation of the original data is then duplicated $k-1$ times, one for each dichotomization. After "blowing up" the data, a standard conditional logit estimation with clustered standard errors is applied to the whole sample. For more details we refer to the paper of Baetschmann et al. (2011).

Finally, we consider the linear fixed effects model that assumes cardinality and makes use of all variation in individuals' ordered responses, while also accounting for non-random personality traits. The ordered response categories $k = 1, \ldots, K$ of y_{it} are interpreted as continuous values of the latent variable $y{*}_{it}$, which lends itself to linear regression methods. Personality traits can be addressed by, for instance, within-transformation of Equation 1, such that α_i cancels out:

$$y^*_{it} - \overline{y}^*_{it} = \beta'(x_{it} - \overline{x}_{it}) + \epsilon_{it} - \overline{\epsilon}_{it} \tag{7}$$

In the following we refer to this estimation strategy as the fixed effects estima-

[1] For example, a data setup of 3,000 individuals with 15 observations over time can take about half an hour computation time.

tor (FE).[2] The FE has the advantage that it is fast and very easy to implement. However, assuming cardinality of ordered responses may be an assumption yielding biased estimates. Nevertheless, as previously discussed, numerous studies have used this approach (e.g., Scheve and Slaughter, 2004; Di Tella et al., 2001, Senik, 2004) and at least in the context of life satisfaction studies, there is some evidence that the associated bias is only moderate (Ferrer-i-Carbonell and Frijters, 2004). Additionally, Greene (1981), Chung and Goldberger (1984) or Deaton and Irish (1984) theoretically show that, under certain distributional assumptions of the explanatory variables, coefficient estimates of limited dependent variable and discrete choice models using OLS can be consistent up to a scalar multiple. If this is also the case for linearly estimated ordered response models with fixed effects, coefficient ratios of OLS estimates should be consistent as well.

A mild alteration to the FE method is proposed in van Praag and Ferrer-i-Carbonell (2008). Their probit adapted OLS estimator (POLS) attempts to cardinalize the data such that it can be applied to simple OLS without the aforementioned problems of the FE estimator and has been used in e.g., Stevenson and Wolfers (2008), Luechinger (2009), Luechinger et al. (2010). The POLS estimator attempts to circumvent violations of the cardinality assumption by first calculating the relative frequencies of the different outcome categories and then putting the frequencies into a standard normal distribution function to obtain a standard normal distributed, "cardinal scaled", and unbounded dependent variable. This variable can be used then for simple (fixed effect) OLS. For more details on this procedure, see Chapter 2.6 in van Praag and Ferrer-i-Carbonell (2008).

Regardless, from a theoretical perspective, assuming cardinality of ordered responses may be unsatisfactory, and our Monte Carlo simulations will show whether this pragmatic approach frequently employed in the life satisfaction literature is justified in a more general setting.

[2] First difference transformation of the model yields equivalent results.

3 Monte Carlo simulation and results

For some of the analysed estimation strategies asymptotic properties have been theoretically proven in the literature, other namely the FCF, IMCLOG and BUC still have unknown asymptotic properties. Regardless, for the applied researcher finite sample properties are important when choosing between different estimation strategies. We therefore perform Monte Carlo simulations for all estimators discussed in Section 2 to provide a guideline for when which estimation strategy is appropriate. We also consider the standard ordered logit without unobserved personality traits controls for comparison. All simulations are performed 1000 times for different sample sizes, ordinal scales, number and distribution of covariates.[3] Our data generating process is designed in line with the standard Monte Carlo simulation literature for panel data (e.g., Honoré and Kyriazidou, 2000; Greene, 2004). The latent variable y_{it}^* is generated by the following model:

$$y_{it}^* = x_{it}\beta + \alpha_i + \epsilon_{it}$$

The individual fixed effect α_i is generated as $\alpha_i = \sqrt{T}\bar{x}_i$. The idiosyncratic error ϵ_{it} is i.i.d. logistically distributed, and the exogenous variables x_{it} are i.i.d. normally distributed. Both error and exogenous variables have the same standard deviation of $\sigma = \pi/\sqrt{3}$. As a robustness check we later consider alternative symmetric and asymmetric distributions of x_{it}.

We define the categories for the discrete dependent variable y_{it} by splitting the generated latent variable y_{it}^* into K even parts. As a result, every category has the same number of observations. To evaluate how the different estimates converge to the true parameters, we focus on the mean of the estimated coefficients, the mean squared error (MSE), and as a more robust performance measure to possible outliers, the median absolute error (MAE). We also compare efficiency measures like the mean of the coefficients' standard errors (S.E.) as well as associated 95 per cent confidence intervals across simulations.

We start with only one exogenous variable x_{it} and set the coefficient to $\beta = 1$. To

[3]We use the statistical software STATA to run our simulations. The corresponding STATA ado-file for the FCF estimator is attached as a additional file in the submission.

compare the asymptotic properties of the estimators under consideration we start with a small panel and subsequently increase the cross-sectional and longitudinal dimension sizes. Table 1 presents estimation results where we fix the longitudinal dimension to $T = 5$ and raise the cross-sectional dimension size from $I = 100$ to $I = 3,000$ while $K = 3$.

In accordance with asymptotic theory, all nonlinear estimators except IMCLOG and FCF and of course the standard ordered logit converge towards the true parameter with growing precision with increasing I. When instead evaluating asymptotic properties over t, as reported in Table 2 [4] we see in our simulations that the coefficient estimates converges towards the true parameter for all nonlinear estimators except the standard ordered logit. Unsurprisingly, the class of linear estimators (FE and POLS) cannot provide consistent estimates of the true parameter due to the the different functional form of the probability function. As a consequence, with only one explanatory variable, the FE and POLS cannot be compared with the other estimators, and we do not report performance measures other than the mean coefficients and standard errors. However, when later including more than one explanatory variable, we will compare the coefficient ratios to reflect on the relative size of coefficients.

In what follows we look at the speed with which convergence of the different nonlinear estimators is achieved and how severe the bias of inconsistent estimators is. Ignoring unobserved individual heterogeneity clearly biases coefficient estimates for all panel data configurations. In Table 1 and Table 2 the means of the simple ordered logit coefficients are always furthest away from the true parameter $\beta = 1$. These simulation results are in line with Ferrer-i-Carbonell and Frijters (2004) who stress the importance of allowing for individual fixed effects.

Comparing the consistent nonlinear models SCLOG and BUC leads to several important insights. First of all, the simple binary coding procedure SCLOG is very sensitive to small sample sizes because it already disregards a large part of the available variation in the dependent variable.[5] For example, with $T = 5$ and

[4] We also perform simple t-tests to compare the means of the respective estimators' coefficients when I and T increase. The differences of the means are statistically significant when starting from small T and small I and become insignificant when both dimension sizes are large.

[5] For our data set with $y_{it} \in \{1, 2, 3\}$ we did the following binary recoding: $y_{it}^n = 1$, if $y_{it} > 2$.

$I = 100$, 40 percent of all observations were ignored because of no variation in the dependent variable. With real survey data and less homogeneous categories, the loss of variation may be even more serious. We therefore recommend not using the SCLOG method in small samples.

Of all estimators the BUC method dominates in terms of consistency and efficiency measures for all panel data configurations. Through all our simulations the mean of estimated parameters is closest to the true value $\beta = 1$ with the lowest MSE and MAE. At the same time the mean standard error of β and the associated confidence interval is smallest. As a first conclusion, these simulations clearly show the asymptotic properties of the estimation methods: Only the SCLOG and BUC estimates can be considered as unbiased while the BUC is most efficient.

We proceed by comparing the set of estimators when including more than one explanatory variable in the model, which is more informative for the applied researcher. Table 3 reports the performance measures for the coefficient with three explanatory variables. In applied research, coefficient ratios are frequently employed to interpret the size of coefficients relative to a baseline effect. In the analysis of individual well-being, for instance, it is common to calculate compensating income variations, i.e., the well-being effect of certain events expressed in percentage changes in income that would generate the same well-being effect (see Winkelmann and Winkelmann, 1998). Accordingly, it is not necessarily the absolute size of coefficients that researchers are interested in, but their ratios.

For the following simulation, we arbitrarily set total number of observations to 18,000 consisting of $I = 3000$ and $T = 6$, a sample structure not uncommon in micro data. We choose $\beta_1 = 1$, $\beta_2 = -3.5$ and $\beta_3 = 7$ as the true data generating parameters so we can also evaluate the correct sign of the parameter estimates as well as their ratios $\beta_2/\beta_1 = -3.5$ and $\beta_3/\beta_1 = 7$.

As previously argued, the coefficients of the linear fixed effects models (FE, POLS) cannot be compared to the ones from nonlinear estimators due to the different scaling. However, as becomes apparent in Table 3 the estimated coefficient ratios of the FE, as well as the ratios of the POLS, are very close to the ratios of the true parameters, i.e., $\widehat{\beta}_2/\widehat{\beta}_1$ is almost exactly -3.5 and $\widehat{\beta}_3/\widehat{\beta}_1$ is nearly 7. At the same time, of all estimators, the MSE and the MAE of the FE and the POLS are smallest.[6]

Of all the nonlinear estimators controlling for unobserved heterogeneity in Table 3, both the BUC and the FCF method outperform the others in terms of unbiasedness and efficiency of coefficient ratio estimates. Compared to the SCLOG and the IMCLOG, the means of the BUC and FCF parameter estimates come closest to the true parameters in conjunction with the smallest standard errors and lowest values for MSE and MAE. In comparison, ignoring unobserved individual heterogeneity by applying the simple ordered logit estimator leads to severely biased coefficient ratios in Table 3. This becomes apparent when looking at the 95 per cent interval of the ordered logit estimates, in which the true parameters are not included, and the large MAE.

We also check the performance of the alternative estimation strategies for different distributions of the explanatory variables. Table 4 shows Monte Carlo simulations for left and right skewed Beta distributions as well as for normal distributions with different first and second moments. In general, when departing from the standard normal distribution the bias of estimated coefficient ratios increases. However, we still infer that the FE, POLS, BUC and FCF deliver coefficient ratio estimates with small bias.

So far we have assumed that the ordinal response variable is fairly aggregated and lies on a three-point scale ($K = 3$). However, various ordinal scales consist of more than three categories. For example, in the U.S. National Survey of Families and Households (NSFH) and the German Socio-Economic Panel (SOEP), information on individual well-being is captured on a seven- and eleven-point scale, respectively. Against this backdrop, we want to test the extent to which the performance of the estimators under consideration varies with respect to the ordinal structure of the dependent variable. Table 5 lists the simulation results for a three-, seven- and eleven-point scale ordered response variable. All simulations are performed with two exogenous variables with the true parameters $\beta_1 = 1$ and $\beta_2 = -2$. The panel data

[6]Furthermore, our simulations for increasing samples sizes (not reported) indicate that the FE and POLS deliver in fact consistent estimates of parameter ratios.

dimensions are $I = 3,000$ and $T = 12$.[7] Interestingly, it seems that the IMCLOG and FCF method respond rather sensitively to the number of ordered categories in the dependent variable. With increasing K the estimated parameters show a sizeable downward bias, although the $\beta_2/\beta_1 = -2$ ratios remain unbiased. This confirms Baetschmann et al. (2011) who have recently shown, that the estimation strategies of Ferrer-i-Carbonell and Frijters (2004) of which IMCLOG can be considered a special case can produce biased parameter estimates. The reason behind is an endogeneity problem of the individual threshold, which is by itself a function of the original ordered variable. In comparison, BUC, and SCLOG are not sensitive with respect to the size of K; there is no significant change in the mean of the parameter estimates, the MSE, MAE or in the the mean standard error. In terms of coefficient ratios, all nonlinear estimates are unbiased as long unobserved personality traits are controlled for, irrespective of K. The same holds for the linear class of estimators FE and POLS.

Summarizing our simulation results, we find the BUC estimator to perform best, that is to deliver unbiased and efficient parameter estimates irrespective of sample size, the underlying distribution of x_{it} and the number of ordinal response categories. In addition, for large samples the SCLOG estimator also performs well and may be even easier to implement.

However, if the researcher is only interested in relative parameters, all of the above estimators deliver unbiased parameter ratios as long as unobserved personality traits are controlled for. This finding also relates to a large theoretical literature that proves that even with misspecified nonlinear models one can obtain consistent coefficient ratio estimates (see e.g., Ruud, 1983; Cramer, 2007 and Wooldridge, 2010).

Furthermore, our Monte Carlo simulations show that to obtain unbiased estimates of parameter ratios one can also employ simple linear estimation allowing for individual fixed effects. This is a generalization of a familiar result derived in the context of happiness studies (e.g., Ferrer-i-Carbonell and Frijters, 2004). The result also relates to Greene (1981), Chung and Goldberger (1984) or Deaton and

[7] To accommodate higher K it is necessary to have more observations per individual. We therefore increase the number of time periods from $T = 6$ in Table 3 to $T = 12$.

Irish (1984) who demonstrate that under certain distributional assumptions one can obtain consistent parameter ratio estimates by applying OLS to discrete choice problems.

4 Conclusion

We compare linear and nonlinear ordered response estimators in terms of consistency and efficiency measures by running Monte Carlo simulations while varying the sample size, the number and distribution of covariates, and the number of ordinal response categories. The estimators under consideration are linear fixed effects, probit adapted OLS, simple ordered logit, and four binary recoded conditional logit estimators that recently have gained popularity in applied research.

Our simulations indicate that first of all it is crucial to control for individual unobserved heterogeneity. Failing to do so adds considerable bias to estimates of parameters and parameter ratios. If the researcher is interested in the absolute size of parameter estimates as such the best choice for estimating ordered response models is the newly developed "Blow-Up and Cluster" estimator of Baetschmann et al. (2011). It delivers most unbiased and most efficient parameter estimates, irrespective of sample size and number of ordinal response categories. The simple conditional logit estimator is an even more basic alternative but only appropriate for large samples.

However, if the researcher is mainly interested in relative effects, i.e. in ratios of parameter estimates, the method of choice is simple: a linear fixed effects model. It essentially delivers the same results as the more elaborate binary recoding scheme of Baetschmann et al. (2011) and is most efficient and much easier to compute.

References

Baetschmann, Gregori, Kevin E. Staub, and Rainer Winkelmann, "Consistent Estimation of the Fixed Effects Ordered Logit Model," Discussion Paper 5443, IZA 2011.

Chamberlain, Gary, "Analysis of Covariance with Qualitative Data," *Review of Economic Studies*, January 1980, *47*, 225–238.

Chung, Ching-Fan and Arthur S. Goldberger, "Proportional Projections in Limited Dependent Variable Models," *Econometrica*, 1984, *52* (2), 531–534.

Clark, Andrew, "Unemployment as a Social Norm: Psychological Evidence from Panel Data," *Journal of Labor Economics*, 2003, *21*, 323–351.

_ , Andreas Knabe, and Steffen Rätzel, "Boon or bane? Others' unemployment, well-being and job insecurity," *Labour Economics*, 2010, *17*, 52–61.

Cramer, J. S., "Robustness of Logit Analysis: Unobserved Heterogeneity and Mis-specified Disturbances," *Oxford Bulletin of Economics and Statistics*, August 2007, *69* (4), 545–555.

Deaton, Angus and Margaret Irish, "Statistical models for zero expenditures in household budgets," *Journal of Public Economics*, 1984, *23* (1-2), 59–80.

Di Tella, Rafael, Robert J. MacCulloch, and Andrew J. Oswald, "Preferences over Inflation and Unemployment: Evidence from Surveys of Happiness," *American Economic Review*, 2001, *91*, 335–341.

Ferrer-i-Carbonell, Ada and Paul Frijters, "How Important is Methodology for the Estimates of the Determinants of Happiness?," *Economic Journal*, 2004, *114*, 641–659.

Frijters, Paul, Ingo Geishecker, John Haisken-De-New, and Michael Shields, "Can the Large Swings in Russian Life Satisfaction be Explained by Ups and Downs in Real Incomes?," *Scandinavian Journal of Economics*, 2006, *108* (3), 433–458.

__ , **John P Haisken-DeNew, and Michael A Shields**, "Money Does Matter! Evidence From Increasing Real Income and Life Satisfaction in East Germany Following Reunification," *American Economic Review*, 2004, *94*, 730–740.

Geishecker, Ingo, "Simultaneity Bias in the Analysis of Perceived Job Insecurity and Subjective Well-Being," *Economics Letters*, 2012, *116* (3), 319–321.

__ , **Maximilian Riedl, and Paul Frijters**, "Offshoring and job loss fears: An econometric analysis of individual perceptions," *Labour Economics*, 2012, *19* (5), 738–747.

Greene, William, "The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects," *Econometrics Journal*, 2004, *7* (1), 98–119.

Greene, William H., "On the Asymptotic Bias of the Ordinary Least Squares Estimator of the Tobit Model," *Econometrica*, 1981, *49* (2), 505–513.

Honoré, Bo E. and Ekaterini Kyriazidou, "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica*, July 2000, *68* (4), 839–874.

Kassenböhmer, Sonja C and John P Haisken-DeNew, "You're Fired! The Causal Negative Effect of Unemployment on Life Satisfaction," *The Economic Journal*, 2009, *119*, 448–462.

Knabe, Andreas and Steffen Rätzel, "Scarring or Scaring? The Psychological Impact of Past Unemployment and Future Unemployment Risk," *Economica*, 2009, *119* (536), 448–462.

Luechinger, Simon, "Valuing Air Quality Using The Life Satisfaction Approach," *Economic Journal*, 2009, *119*, 482–515.

__ , **Stephan Meier, and Alois Stutzer**, "Why Does Unemployment Hurt the Employed? Evidence from the Life Satisfaction Gap Between the Public and the Private Sector," *Journal of Human Resources*, 2010, *45* (4), 998–1045.

Ruud, Paul A., "Sufficient Conditions for the Consistency of Maximum Likelihood Estimation Despite Misspecification of Distribution in Multinomial Discrete Choice Models," *Econometrica*, January 1983, *51* (1), 225–228.

Scheve, Kenneth and Metthew J Slaughter, "Economic Insecurity and the Globalization of Production," *American Journal of Political Science*, 2004, *48* (4), 662–674.

Senik, Claudia, "When information dominates comparison: Learning from Russian subjective panel data.," *Journal of Public Economics*, 2004, *88*, 9–10.

Stevenson, Betsey and Justin Wolfers, "Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox," *Brookings Papers on Economic Activity*, 2008, *2008* (1), 1–87.

van Praag, Bernard and Ada Ferrer-i-Carbonell, *Happiness Quantified: A Satisfaction Calculus Approach*, Oxford University Press, 2008.

Winkelmann, Liliana and Rainer Winkelmann, "Why Are the Unemployed So Unhappy? Evidence from Panel Data.," *Economica*, 1998, *65*, 1–15.

Wooldridge, Jeffrey M., *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., The MIT Press, 2010.

Appendix

Loglikelihood equation of the conditional logit model:

$$\ln L_{ik} = \sum_{t=1}^{T} D_{itk} x_{it} \beta - \ln \sum_{S\left(\sum_{t=1}^{T} D_{itk}\right)} e^{\sum_{t=1}^{T} D_{itk} x_{it} \beta}$$

Gradient function of the conditional logit model:

$$\frac{\partial \ln L_{ik}}{\partial \beta} = \sum_{t=1}^{T} D_{itk} x_{it} - \frac{\sum_{S\left(\sum_{t=1}^{T} D_{itk}\right)} \left(\sum_{t=1}^{T} D_{itk} x_{it}\right) e^{\sum_{t=1}^{T} D_{itk} x_{it} \beta}}{\sum_{S\left(\sum_{t=1}^{T} D_{itk}\right)} e^{\sum_{t=1}^{T} D_{itk} x_{it} \beta}}$$

Hessian function of the conditional logit model:

$$H = \frac{\partial^2 \ln L_{ik}}{\partial \beta^2}$$

$$H = \frac{\left(\sum_{S\left(\sum_{t=1}^{T} D_{itk}\right)} \left(\sum_{t=1}^{T} D_{itk} x_{it}\right) e^{\sum_{t=1}^{T} D_{itk} x_{it} \beta}\right) \left(\sum_{S\left(\sum_{t=1}^{T} D_{itk}\right)} \left(\sum_{t=1}^{T} D_{itk} x_{it}\right) e^{\sum_{t=1}^{T} D_{itk} x_{it} \beta}\right)}{\left(\sum_{S\left(\sum_{t=1}^{T} D_{itk}\right)} e^{\sum_{t=1}^{T} D_{itk} x_{it} \beta}\right)^2}$$

$$- \frac{\left[\sum_{S\left(\sum_{t=1}^{T} D_{itk}\right)} \left(\sum_{t=1}^{T} D_{itk} x_{it}\right) \left(\sum_{t=1}^{T} D_{itk} x_{it}\right) e^{\sum_{t=1}^{T} D_{itk} x_{it} \beta}\right] \sum_{S\left(\sum_{t=1}^{T} D_{itk}\right)} e^{\sum_{t=1}^{T} D_{itk} x_{it} \beta}}{\left(\sum_{S\left(\sum_{t=1}^{T} D_{itk}\right)} e^{\sum_{t=1}^{T} D_{itk} x_{it} \beta}\right)^2}$$

$$= A * A - \frac{\sum_{S\left(\sum_{t=1}^{T} D_{itk}\right)} \left(\sum_{t=1}^{T} D_{itk} x_{it}\right) \left(\sum_{t=1}^{T} D_{itk} x_{it}\right) e^{\sum_{t=1}^{T} D_{itk} x_{it} \beta}}{\sum_{S\left(\sum_{t=1}^{T} D_{itk}\right)} e^{\sum_{t=1}^{T} D_{itk} x_{it} \beta}}$$

With $A = \frac{\sum_{S\left(\sum_{t=1}^{T} D_{itk}\right)} \left(\sum_{t=1}^{T} D_{itk} x_{it}\right) e^{\sum_{t=1}^{T} D_{itk} x_{it} \beta}}{\sum_{S\left(\sum_{t=1}^{T} D_{itk}\right)} e^{\sum_{t=1}^{T} D_{itk} x_{it} \beta}}$ corresponding to the second term of the gradient function.

Table 1: Monte Carlo simulation results for $K = 3$, $T = 5$

	$\boldsymbol{\beta = 1}$					
	Mean	**S.E.**	**MSE**	**MAE**	**95% Interval**	
I = 100						
FE OLS	0.20526	0.01416				
POLS	0.22324	0.01540				
ordered logit	1.03623	0.06999	0.00671	0.05192	0.89769	1.20035
SCLOG	1.01140	0.14563	0.02093	0.09346	0.76928	1.32656
FCF	0.98475	0.11516	0.02538	0.08207	0.78009	1.24946
IMCLOG	0.98594	0.11772	0.01531	0.07952	0.78337	1.25336
BUC	1.00708	0.10682	0.01238	0.06708	0.80646	1.25820
I = 500						
FE OLS	0.20555	0.00632				
POLS	0.22355	0.00687				
ordered logit	1.03423	0.03124	0.00215	0.03322	0.97628	1.10259
SCLOG	1.00433	0.06419	0.00446	0.04419	0.88182	1.14270
FCF	0.97926	0.05102	0.00314	0.03892	0.88404	1.08702
IMCLOG	0.98090	0.05218	0.00310	0.03846	0.88357	1.08989
BUC	1.00330	0.04780	0.00242	0.03409	0.91493	1.11028
I = 1000						
FE OLS	0.20477	0.00446				
POLS	0.22270	0.00485				
ordered logit	1.03298	0.02206	0.00163	0.00840	0.98943	1.07798
SCLOG	1.00183	0.04529	0.00225	0.03235	0.91389	1.09804
FCF	0.97711	0.03600	0.00193	0.03270	0.91003	1.05655
IMCLOG	0.97921	0.03684	0.00191	0.03224	0.90987	1.06044
BUC	1.00080	0.03390	0.00124	0.02508	0.93906	1.07563
I = 3000						
FE OLS	0.20492	0.00258				
POLS	0.22286	0.00280				
ordered logit	1.03253	0.01275	0.00122	0.00835	1.00751	1.05610
SCLOG	0.99857	0.02603	0.00064	0.01767	0.95139	1.05082
FCF	0.97514	0.02073	0.00103	0.02561	0.93725	1.01506
IMCLOG	0.97694	0.02121	0.00096	0.02343	0.93794	1.01708
BUC	0.99912	0.01953	0.00037	0.01362	0.96365	1.03747

Note: All simulations were performed 1000 times.

Table 2: Monte Carlo simulation results for $K = 3$, $I = 1000$

	$\boldsymbol{\beta = 1}$					
	Mean	**S.E.**	**MSE**	**MAE**	**95% Interval**	
T = 3						
FE OLS	0.19821	0.00620				
POLS	0.21557	0.00674				
ordered logit	1.17846	0.03299	0.03296	0.17729	1.11813	1.24220
SCLOG	1.00044	0.07233	0.00544	0.04885	0.86150	1.15537
FCF	0.96420	0.05460	0.00424	0.04818	0.86095	1.07516
IMCLOG	0.97921	0.05570	0.00427	0.04775	0.85826	1.08004
BUC	0.99988	0.05383	0.00286	0.03576	0.89692	1.10606
T = 5						
FE OLS	0.20555	0.00632				
POLS	0.22355	0.00687				
ordered logit	1.03423	0.03124	0.00215	0.03322	0.97628	1.10259
SCLOG	1.00433	0.06419	0.00446	0.04419	0.88182	1.14270
FCF	0.97926	0.05102	0.00314	0.03892	0.88404	1.08702
IMCLOG	0.98090	0.05218	0.00310	0.03846	0.88357	1.08989
BUC	1.00330	0.04780	0.00242	0.03409	0.91493	1.11028
T = 10						
FE OLS	0.21262	0.00304				
POLS	0.23124	0.00330				
ordered logit	0.90834	0.01379	0.00859	0.09168	0.88139	0.93454
SCLOG	0.99986	0.02741	0.00074	0.01834	0.94901	1.05545
FCF	0.98763	0.02310	0.00070	0.01823	0.94010	1.03268
IMCLOG	0.98774	0.02343	0.00071	0.01835	0.93772	1.03382
BUC	0.99917	0.02063	0.00043	0.01415	0.95767	1.04011
T = 15						
FE OLS	0.21602	0.00246				
POLS	0.23493	0.00267				
ordered logit	0.85903	0.01103	0.01999	0.14065	0.83816	0.88044
SCLOG	1.00004	0.02135	0.00050	0.01496	0.95837	1.04591
FCF	0.99102	0.01839	0.00045	0.01476	0.95502	1.02925
IMCLOG	0.99116	0.01858	0.00044	0.01507	0.95451	1.02742
BUC	0.99956	0.01614	0.00028	0.01139	0.96684	1.03459

Note: All simulations were performed 1000 times.

Table 3: Monte Carlo simulation results for $K = 3, I = 3000, T = 6$

$\boldsymbol{\beta_2/\beta_1 = -3.5}$

	Mean	**MSE**	**MAE**	**95% Interval**	
FE OLS	-3.50254	0.01688	0.09059	-3.77135	-3.26882
POLS	-3.50255	0.01688	0.09047	-3.77223	-3.26900
ordered logit	-2.96128	0.29334	0.54159	-3.07308	-2.85501
SCLOG	-3.50610	0.03664	0.12148	-3.92625	-3.16180
FCF	-3.50022	0.01837	0.08657	-3.78670	-3.23716
IMCLOG	-3.50289	0.02615	0.10578	-3.85568	-3.21248
BUC	-3.49951	0.01807	0.08461	-3.78585	-3.24382

$\boldsymbol{\beta_3/\beta_1 = 7}$

	Mean	**MSE**	**MAE**	**95% Interval**	
FE OLS	7.00921	0.06358	0.17499	6.55400	7.54091
POLS	7.00924	0.06358	0.17470	6.55289	7.54045
ordered logit	6.28513	0.52405	0.71975	6.07888	6.51445
SCLOG	7.01410	0.13999	0.24292	6.33485	7.85700
FCF	7.00274	0.06840	0.17344	6.51613	7.55918
IMCLOG	7.00530	0.09721	0.19735	6.43982	7.68614
BUC	7.00133	0.06733	0.17449	6.52065	7.54664

Note: All simulations were performed 1000 times.

Table 4: Monte Carlo simulation results for different distributions of the explanatory variables

$x_1 \sim \text{Beta}(1,5), x_2 \sim \text{Beta}(2,2), x_3 \sim \text{Beta}(5,1)$

	$\beta_2/\beta_1 = -3.5$		$\beta_3/\beta_1 = 7$		$\beta_3/\beta_2 = -2$	
Method	**Mean**	**MSE**	**Mean**	**MSE**	**Mean**	**MSE**
FE OLS	-3.69695	0.78830	7.01602	2.67229	-1.89997	0.01798
POLS	-3.69685	0.78805	7.01293	2.66892	-1.89918	0.01812
SCLOG	-3.71210	1.26982	7.42891	5.07450	-2.00496	0.01442
FCF	-3.68803	1.19382	7.39570	4.74821	-2.00864	0.01240
IMCLOG	-3.69735	1.09772	7.39250	4.29584	-2.00282	0.01260
BUC	-3.63606	0.76707	7.27752	3.05166	-2.00384	0.00940

$x_1 \sim \text{Normal}(0,1), x_2 \sim \text{Normal}(5,10), x_3 \sim \text{Normal}(2,0.1)$

	$\beta_2/\beta_1 = -3.5$		$\beta_3/\beta_1 = 7$		$\beta_3/\beta_2 = -2$	
Method	**Mean**	**MSE**	**Mean**	**MSE**	**Mean**	**MSE**
FE OLS	-3.74551	1.32221	7.42532	11.96346	-1.99623	0.51903
POLS	-3.74530	1.32103	7.42451	11.95713	-1.99613	0.51894
SCLOG	-4.37796	35.83729	8.19181	256.44650	-1.89413	1.19050
FCF	-3.81028	1.91946	7.39687	17.39827	-1.93452	0.57558
IMCLOG	-3.98390	17.75278	7.49777	96.14915	-1.93663	0.94279
BUC	-3.80662	1.91339	7.36381	17.26725	-1.92684	0.57119

$x_1 \sim \text{Normal}(0,1), x_2 \sim \text{Normal}(0,2), x_3 \sim \text{Normal}(0,3)$

	$\beta_2/\beta_1 = -3.5$		$\beta_3/\beta_1 = 7$		$\beta_3/\beta_2 = -2$	
Method	**Mean**	**MSE**	**Mean**	**MSE**	**Mean**	**MSE**
FE OLS	-3.59997	0.51341	7.20573	2.06420	-2.00254	0.00271
POLS	-3.59987	0.51328	7.20553	2.06373	-2.00255	0.00271
SCLOG	-3.74290	1.34155	7.48449	5.29469	-2.00199	0.00547
FCF	-3.63117	0.54875	7.26366	2.19126	-2.00103	0.00273
IMCLOG	-3.68796	0.82737	7.37118	3.29663	-1.99967	0.00371
BUC	-3.62824	0.54313	7.25735	2.16254	-2.00097	0.00270

Monte Carlo simulation results for $I = 1000, T = 6, K = 3$.
All simulations were performed 1000 times.

Table 5: Monte Carlo simulation results for $I = 3000, T = 12$

	$\beta_1 = 1$					
	Mean	**S.E.**	**MSE**	**MAE**	**95 % Interval**	
K = 3						
FE OLS	0.16148	0.00248				
POLS	0.17563	0.00270				
ordered logit	0.92901	0.01586	0.00529	0.07188	0.89849	0.96411
SCLOG	0.99994	0.02903	0.00090	0.01946	0.94103	1.06088
FCF	0.99330	0.02656	0.00074	0.01819	0.94225	1.04474
IMCLOG	0.98699	0.02651	0.00094	0.02148	0.93630	1.04430
BUC	0.99976	0.02110	0.00045	0.01412	0.96138	1.04181
K = 7						
FE OLS	0.42420	0.00500				
POLS	0.20693	0.00235				
ordered logit	0.92531	0.01208	0.00572	0.07466	0.90169	0.94856
SCLOG	0.99972	0.02784	0.00075	0.01826	0.94528	1.05358
FCF	0.95820	0.02539	0.00239	0.04203	0.90878	1.00931
IMCLOG	0.99042	0.02635	0.00080	0.01979	0.93930	1.04363
BUC	0.99932	0.01645	0.00026	0.01100	0.96783	1.03000
K = 11						
FE OLS	0.67823	0.00761				
POLS	0.21344	0.00225				
ordered logit	0.92525	0.01138	0.00573	0.07483	0.90294	0.94886
SCLOG	0.99906	0.02766	0.00078	0.01933	0.94290	1.05466
FCF	0.93817	0.02481	0.00449	0.06286	0.89028	0.99085
IMCLOG	0.98989	0.02628	0.00081	0.01933	0.93778	1.03975
BUC	0.99942	0.01575	0.00027	0.01090	0.96750	1.03096

	$\beta_2 = -2$					
	Mean	**S.E.**	**MSE**	**MAE**	**95% Interval**	
K = 3						
FE OLS	-0.32337	0.00248				
POLS	-0.35169	0.00269				
ordered logit	-1.50856	0.02589	0.24219	0.49096	-1.56013	-1.45887
SCLOG	-2.00233	0.04700	0.00242	0.03288	-2.10216	-1.90702
FCF	-1.98814	0.04479	0.00226	0.03063	-2.07914	-1.89968
IMCLOG	-1.97595	0.04291	0.00265	0.03646	-2.06764	-1.89027
BUC	-2.00110	0.03390	0.00124	0.02316	-2.07578	-1.93391
K = 7						
FE OLS	-0.84926	0.00500				
POLS	-0.41421	0.00235				
ordered logit	-1.50239	0.02181	0.24808	0.49766	-1.54547	-1.46093
SCLOG	-2.00188	0.04511	0.00197	0.02853	-2.09208	-1.91451
FCF	-1.91857	0.04403	0.00865	0.08329	-2.01559	-1.83207
IMCLOG	-1.98281	0.04264	0.00213	0.03225	-2.06524	-1.89991
BUC	-1.99984	0.02595	0.00070	0.01730	-2.05347	-1.94883
K = 11						
FE OLS	-1.35736	0.00761				
POLS	-0.42709	0.00225				
ordered logit	-1.50084	0.02107	0.24960	0.49944	-1.54128	-1.45818
SCLOG	-2.00034	0.04480	0.00201	0.02971	-2.08792	-1.91651
FCF	-1.87649	0.04313	0.01711	0.12460	-1.96627	-1.79364
IMCLOG	-1.98237	0.04254	0.00219	0.03074	-2.06710	-1.89292
BUC	-1.99987	0.02479	0.00067	0.01746	-2.05021	-1.94826

continued on next page...

Table 5: ... continued

	$\beta_2/\beta_1 = -2$				
	Mean	**MSE**	**MAE**	**95% Interval**	
K = 3					
FE OLS	-2.00295	0.00114	0.02251	-2.06771	-1.93409
POLS	-2.00295	0.00114	0.02261	-2.06770	-1.93408
ordered logit	-1.62404	0.14204	0.37591	-1.67641	-1.57123
SCLOG	-2.00323	0.00201	0.02935	-2.09963	-1.91360
FCF	-2.00207	0.00137	0.02585	-2.07552	-1.93191
IMCLOG	-2.00268	0.00175	0.02665	-2.08529	-1.92115
BUC	-2.00198	0.00106	0.02066	-2.06850	-1.93866
K = 7					
FE OLS	-2.00229	0.00074	0.01871	-2.05785	-1.95103
POLS	-2.00200	0.00067	0.01766	-2.05361	-1.95338
ordered logit	-1.62376	0.14202	0.37563	-1.66442	-1.58330
SCLOG	-2.00312	0.00172	0.02757	-2.08770	-1.92411
FCF	-2.00270	0.00117	0.02266	-2.07857	-1.94011
IMCLOG	-2.00266	0.00168	0.02766	-2.08324	-1.91922
BUC	-2.00146	0.00067	0.01700	-2.05403	-1.95221
K = 11					
FE OLS	-2.00159	0.00070	0.01795	-2.05382	-1.95094
POLS	-2.00119	0.00059	0.01666	-2.04767	-1.95443
ordered logit	-1.62222	0.14322	0.37818	-1.66620	-1.58054
SCLOG	-2.00295	0.00183	0.02941	-2.09252	-1.92420
FCF	-2.00071	0.00130	0.02393	-2.07339	-1.93461
IMCLOG	-2.00328	0.00170	0.02745	-2.08487	-1.92772
BUC	-2.00130	0.00068	0.01717	-2.05481	-1.95115

Note: All simulations were performed 1000 times.